

[ジェンダー,身長,体重]のデータ

```
In [1]: import numpy as np
import pandas as pd
```

```
In [2]: nRowsRead = None # specify 'None' if want to read whole file
# weight-height.csv has 10000 rows in reality, but we are only loading/previewing the first 1000 rows
df1 = pd.read_csv('data_h_w.csv', delimiter=',', nrows = nRowsRead)
df1.dataframeName = 'gender-weight-height'
nRow, nCol = df1.shape
print(f'There are {nRow} rows and {nCol} columns')
```

There are 1000 rows and 4 columns

```
In [3]: df1.head(5)
```

```
Out[3]:
```

	Unnamed: 0	Gender	Height	Weight
0	9953	Female	61.278349	132.534229
1	3850	Male	70.454991	190.824938
2	4962	Male	70.144763	194.831396
3	3886	Male	69.231469	173.829606
4	5437	Female	68.759636	153.689739

```
In [6]: df1=df1.drop('Unnamed: 0',axis=1)
```

```
In [7]: df1.tail()
```

```
Out[7]:
```

	Gender	Height	Weight
995	Male	72.796730	196.504136
996	Female	64.258342	140.236766
997	Male	67.903612	177.352590
998	Male	71.189384	202.782736
999	Male	66.631041	170.421149

```
In [8]: df1.describe(include='all')
```

```
Out[8]:
```

	Gender	Height	Weight
count	1000	1000.000000	1000.000000
unique	2	NaN	NaN
top	Male	NaN	NaN
freq	530	NaN	NaN
mean	NaN	66.475941	162.186259
std	NaN	3.981159	33.052238
min	NaN	55.668202	68.982530
25%	NaN	63.472736	137.364286
50%	NaN	66.438081	163.870956
75%	NaN	69.563558	188.352381
max	NaN	78.095867	255.690835

```
In [9]: df2 = df1.sample(frac=0.1,random_state=1)
df2.describe()
```

```
Out[9]:
```

	Height	Weight
count	100.000000	100.000000
mean	67.263896	168.845554
std	3.705571	31.735864
min	60.955487	105.880581
25%	63.827713	139.702536
50%	67.382045	176.407739
75%	70.158781	195.057177

	Height	Weight
max	75.259457	228.800516

In [10]: df2.head()

	Gender	Height	Weight
507	Male	63.724727	165.979417
818	Female	62.968382	124.773903
452	Male	70.643193	198.491168
368	Male	70.035068	181.324779
242	Male	75.259457	227.437138

In [11]: df2.tail()

	Gender	Height	Weight
671	Male	68.122768	196.414415
559	Female	61.705576	126.787299
593	Male	70.052511	194.974116
258	Male	71.213913	195.923416
154	Female	66.199283	180.698317

In [26]:

In [24]:

In [27]: df2.head()

	Gender	Height	Weight	Height(cm)	Weight(kg)
9953	Female	61.278349	132.534229	156	60.1
3850	Male	70.454991	190.824938	179	86.6
4962	Male	70.144763	194.831396	178	88.4
3886	Male	69.231469	173.829606	176	78.8
5437	Female	68.759636	153.689739	175	69.7

In [28]:

	Height	Weight	Height(cm)	Weight(kg)
count	1000.000000	1000.000000	1000.000000	1000.000000
mean	66.475941	162.186259	168.840000	73.56680
std	3.981159	33.052238	10.138994	14.99224
min	55.668202	68.982530	141.000000	31.30000
25%	63.472736	137.364286	161.000000	62.35000
50%	66.438081	163.870956	169.000000	74.30000
75%	69.563558	188.352381	177.000000	85.42500
max	78.095867	255.690835	198.000000	116.00000

In [35]:

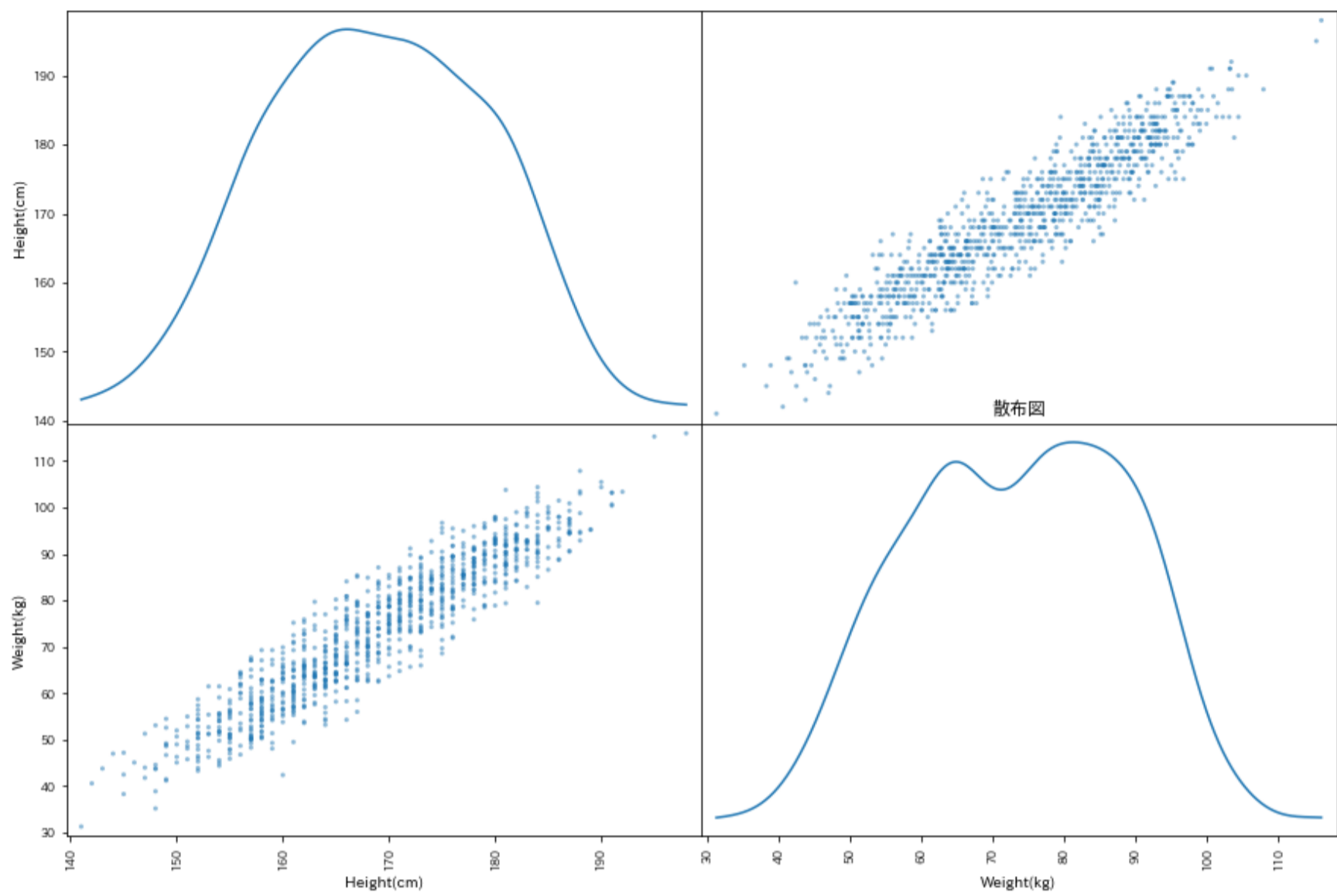
In [36]: df2.head()

	Gender	Height	Weight	Height(cm)	Weight(kg)	BMI
9953	Female	61.278349	132.534229	156	60.1	24.7
3850	Male	70.454991	190.824938	179	86.6	27.0
4962	Male	70.144763	194.831396	178	88.4	27.9
3886	Male	69.231469	173.829606	176	78.8	25.4
5437	Female	68.759636	153.689739	175	69.7	22.8

```
In [51]: %matplotlib inline
```

```
In [54]: import japanize_matplotlib
import matplotlib.pyplot as plt

ax = pd.plotting.scatter_matrix(df2[['Height(cm)', 'Weight(kg)']], figsize=[15, 10], diagonal='kde')
```



```
In [49]: df2[['Height(cm)', 'Weight(kg)']].head()
```

```
Out[49]:
```

	Height(cm)	Weight(kg)
9953	156	60.1
3850	179	86.6
4962	178	88.4
3886	176	78.8
5437	175	69.7

```
In [ ]:
```